Hybrid Image Representation Methods for Automatic Image Annotation: A Survey

Hafidha Bouyerbou[#1], Saliha Oukid [#2], Nadjia Benblidia[#3], Kamal Bechkoum[*4],

[#]*LRDSI Laboratory, Computer Science Department, Saad Dahlab Blida University*
Blida, Algeria

[1]hafidha.bouyerbou@gmail.com

[2]osalyha@yahoo.com

[3]benblidia@yahoo.com

[*]*School of Science and Technology, University of Northampton*
Northampton, United Kingdom

[4]kamal.bechkoum@northampton.ac.uk

***Abstract*— In most automatic image annotation systems, images are represented with low level features using either global methods or local methods. In global methods, the entire image is used as a unit. Local methods divide images into blocks where fixed-size sub-image blocks are adopted as sub-units; or into regions by using segmented regions as sub-units in images. In contrast to typical automatic image annotation methods that use either global or local features exclusively, several recent methods have considered incorporating the two kinds of information, and believe that the combination of the two levels of features is beneficial in annotating images. In this paper, we provide a survey on automatic image annotation techniques according to one aspect: feature extraction, and, in order to complement existing surveys in literature, we focus on the emerging image annotation methods: *hybrid methods* that combine both global and local features for image representation.**

***Keywords*— Image annotation; global features; local features; hybrid methods; feature extraction; image representation.**

I. Introduction

A large amount of research has been conducted on image retrieval (IR) over the last few years. These research efforts can be divided into three broad areas based on the type of approach used [1]. The first approach is the traditional annotation. In this approach, images are annotated manually by humans and then retrieved in the same way as text documents [2-4]. The second approach focuses on content based image retrieval (CBIR), whereby images are automatically indexed and retrieved with low level content features such as colour, shape and texture [5-7]. The third approach is the automatic image annotation (AIA) whereby semantic concept models are learned automatically from large number of image samples; concept models are used to label new images so that images can be retrieved in the same way as text documents.

In general, the AIA is a two-step approach: 1) image component decomposition and representation: by decomposing an image into a collection of sub-units, which could be segmented regions, equal-size blocks or an entire image; and modeling each content unit based on a feature representation scheme; and 2) content classification: by computing the associations between unit representations and textual concepts; in this stage, higher level semantic can be learned from image samples.

For step 1, in most AIA systems, images are represented by either global features where the entire image is used as a unit as in [8-12] and in recent works [13-16]; or block-based local features where fixed-size sub-image blocks are adopted as sub-units for an image as in [17-23]; or region-based local features, by using segmented regions as sub-units in images as in [24-30].

In contrast to typical AIA methods which use either global or block/regional features exclusively, several recent methods have considered incorporating the two kinds of information, and believe that the combination of these two types of features is beneficial in annotating images.

There are several surveys on broad AIA research in literature [1], [31-33] and [55] that deal mainly with the aspect of the semantic learning/annotation and categorize the AIA techniques according to this aspect. However, none of them, to our knowledge, gives sufficient attention to another important issue in AIA namely, *features extraction*.

In this paper, we focus our review on feature extraction for AIA. Specifically, and in order to complement existing surveys in literature, we focus the discussion on the emerging image annotation methods that are *hybrid methods* using both global and local features for the annotation task.

In addition to these introductory notes the paper is organised around a further six sections. Section 2 deals with image representation. Section 3 and 4 are dedicated to several global-feature-based and local-feature-based image annotation techniques, respectively. In section 5, several promising techniques using hybrid image representation methods for image annotation are discussed. Section 6 provides a summary of the survey and some concluding remarks.

II. Image Representation

Image representation is the process of generating descriptions that represent the visual content of images in a certain manner, normally in the form of one or more features [34]. A feature is a function of one or more measurements, which specifies some quantifiable property of an object and quantifies its significant characteristics [35]. Choras [35] classifies the features as follows:

1) General features: Application-independent features, such as color, shape, and texture. They can be further divided into:

- Pixel-level features: the computed features at each single pixel, e.g. color, location.
- Local features: the computed features over the segmented regions or blocks obtained by the subdivision of the image.
- Global features: the computed features over the entire image or the regular sub-area of an image.

2) Domain-specific features: Application-dependent features; they are often a synthesis of low-level features for a specific domain such as human faces, fingerprints, and conceptual features.

On the other hand, all features can be coarsely classified into low-level features and high-level features. While low-level features can be extracted directly from the original images, high-level feature extraction must be based on low-level features [36]. In image classification and retrieval, images are represented using low level features [1].

III. GLOBAL-FEATURE-BASED IMAGE ANNOTATION

The commonly used feature representation is based on a global feature set extracted from images. The global features provide the global distribution of visual topics over an image. Global image features have been widely used in image annotation. In the following, several image annotation approaches based on global feature are reviewed.

Chapelle et al. [9] generalize Support Vector Machines (SVMs) for image classification problems where the only features are high dimensional histograms, and choose to apply them on the global HSV (Hue Saturation Value) colour histograms.

Vailaya et al. [11] use Bayesian classifiers on the color and edge direction histograms to classify vacation photographs into a hierarchy of high-level classes. At the first level, images are classified as indoor or outdoor, the outdoor images are then classified as city or landscape, and finally, a subset of landscape images is further classified into sunset, forest, and mountain classes.

Yavlinsky et al. [12] use non-parametric models of distributions of image features, they present a framework for automated image annotation based on non-parametric density estimation and employ global colour and texture distributions. They use the Earth Mover's Distance (EMD) kernel that only uses global colour information and report results on subsets of two photographic libraries: the Corel Photo Archive and the Getty Image Archive.

Zhang et al. [13], Makadia et al. [14], Babenko et al. [15] and Guillaumin et al. [16] directly transfer labels from training images to test images with global image similarities using a weighted nearest neighbor approach. For example, Makadia et al. [14] extract global color and texture as features; calculate image similarity as the average distance using these features; and the keywords are obtained from the nearest neighbours with the least distance.

IV. LOCAL-FEATURE-BASED IMAGE ANNOTATION

Unlike global features, local features are based on a subset of the image, usually in the neighbourhood of a given point. Local methods are being increasingly used. A number of block-based and region-based methods are reviewed bellow.

A. Block-based image annotation

The simplest way to extract block-based features is to roughly segment images into a fixed number of sub-blocks as shown in Fig. 1. Visual features are then extracted from these blocks. In the following, several image annotation approaches based on block features are reviewed.

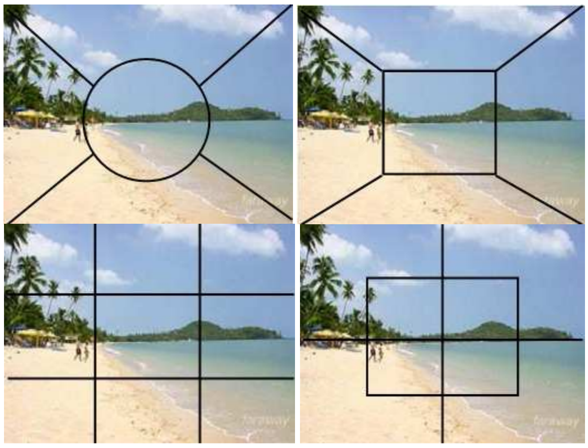


Fig. 1 Examples of block-based division [55]

Gorkani and Picard [17] investigate a measure of dominant perceived orientation. They divide the image into 16 non-overlapping equal-sized blocks. The image is then classified as city/suburb according to the majority orientations of the blocks.

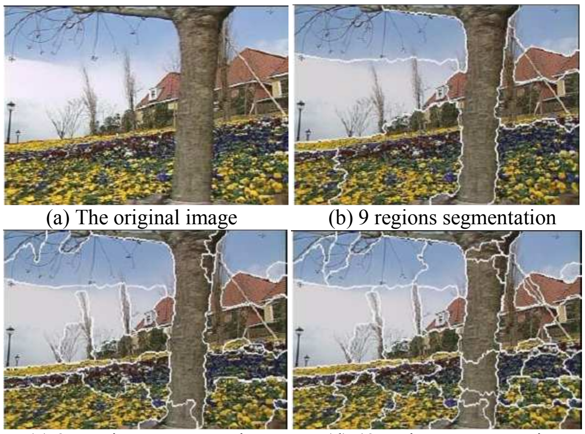
Szummer and Picard [18] first segment each image into a fixed number of blocks; colour and texture features of each block are extracted. Then, a k-NN (K Nearest Neighbor) classifier is designed to classify the colour and texture features of each block into indoor and outdoor categories individually. The final output is based on the blocks of an image which have the highest vote for one of the indoor and outdoor

Serrano et al. [21] use SVMs to classify colour and texture features of 16 blocks per image into indoor and outdoor classes individually. Zhang and Ma [22] propose a block-feature-based multi-class SVM. For image annotation, each image is segmented into five fixed-size blocks instead of time-consuming object segmentation.

Yi and Tang [23] first divide the whole image into different sizes of blocks and generate suitable visual words. Learning is based on the Probabilistic Latent Semantic Analysis (PLSA) by given a set of image blocks for each semantic concept as training data. Finally, the classification of the images is carried out by combining all the image blocks in every block size.

B. Region-based image annotation

The second method for local feature representation is to divide the image into homogenous regions/objects or edges/boundaries using segmentation algorithms as shown in Fig. 2. In the following, several image annotation approaches based on region features, are reviewed.



(a) The original image (b) 9 regions segmentation

(c) 27 regions segmentation (d) 49 regions segmentation

Fig. 2 Examples of region segmentation [55]

Smith and Li [24] propose composite region template descriptor matrix on the spatial orderings of regions to classify image regions into ten categories. Barnard and Forsyth [25] adopt the hierarchical aspect clustering model for image annotation, on semantically meaningful regions to generate words.

Duygulu et al. [26] propose a model of object recognition as machine translation for image annotation. They use Normalized Cuts (N-Cuts) segmentation algorithm to segment image into regions, in which a clustered region "blob" corresponds to a visual vocabulary.

Blei and Jordan [27] describe three hierarchical probabilistic mixture models for a database of annotated images, culminating in correspondence latent Dirichlet allocation (Corr-LDA), a model that finds conditional relationships between latent variable representations of sets of image regions and sets of words, and demonstrate its use in automatic image annotation, automatic region annotation, and text-based image retrieval.

Yang et al. [28] use Multiple Instance Learning (MIL) to learn the correspondence between image regions and words. Tang and Lewis [29] propose to realize automatic region-based image annotation through a training image feature space.

V. HYBRID METHODS

Based on the above, each of the related studies only considers one of the three feature representation methods for automatic image annotation namely, global, block-based, and region-based features.

The combination of these features was considered for face recognition as in [38], [39], [40], [41], [42], and object detection as in [43] and [44]. Promising recent methods have considered incorporating the combination of these features for AIA. In the following, we discuss the emerging image annotation methods that are hybrid methods using both global and local features for the annotation task.

Wang et al. [45] present an approach that combines global, regional, and contextual features by means of an extended Cross-Media Relevance Model (CMRM). They incorporate the three kind of information by estimating their joint probability. The global features are described as a distribution vector of visual topics and model and both global and textual context are learned by a PLSA from the training data. Wang et al. partition an image by a regular grid and take it as an unordered set of image patches. Then extract a 128-D SIFT (Scale Invariant Feature Transform) descriptor and vector-quantize each image patch by clustering a subset of patches from the training images and apply PLSA to learn a set of visual topics. For the region features, Wang et al. uses JSEG algorithm to segment each image into regions and each region is represented by a feature vector including relative region area, color moment feature, shape descriptor and color correlogram feature.

The original CMRM annotates an image using only the regional features. The extended model uses both the regional features and the global features to annotate a test image by estimating the joint probability of a learned textual context, its visual blobs obtained by image segmentation, and the visual topic distribution.

The experiments on Corel image data set show that the proposed approach can yield better annotations than the original CMRM, especially when the test images are associated with multiple labels; and that the annotations have better coherent semantic due to the combination of textual, context, global and regional appearances for the extended CMRM. However, different features have different contribution to a specific word; there was no investigation of these features for image annotation especially for specific categories.

Sarin and Kameyame [46] propose to explore the combination of different visual features at global, local and scene levels including global and local color, texture, and gist of the scene. They extract first the features at image level as well as locally at the Region Of Interest (ROI) level. Then they combine the distance of image equally and use K-NN method for label transfer. Sarin and Kameyame extract three global color histograms RGB, LAB and HSV and the two wavelet textures Haar and Gabor, they compute the colour histogram of the saliency regions for the three color spaces (e.g. RGB, LAB and HSV) and calculate the gist descriptors of two variants of the original image, the first variant is the resized version (256 x 256) and the second one is the square size of the center of the image, and use two distance metric namely, KL-divergence (Kullback-Leibler divergence) as distance metric for LAB and LAB-Saliency and L1 for the other features.

The MIR Flickr 25000 is used in the evaluation and a random combination is applied among all the features and found that the full combination (Color + Texture + Color Saliency + Gist) gives better results in both precision and recall. More importantly, the selective combination of HSV, Haar, HSV saliency, GIST 256 and GIST center gives the best results. Furthermore, they get the best results for labels at k=40. However, in the use of this combination and the selected number k, the evaluation per concept showed that the result is not as good, and some concepts are not selected at all since the method does not provide probability of each annotated concept, they simply give 1 for the detected concept and 0 for the undetected concept.

Tsai and Lin [47] compare the combinations of global, block-based and region-based features by using a standard classifier (i.e. k-NN). For block-based feature, each image is segmented into five 64 × 64 overlapping sub-images as blocks and for region-based feature, the N-Cuts algorithm is used for region segmentation, in which 5 regions of each image are segmented. They consider only color and texture features where the HSV colour space is extracted for colour representation and four levels of Daubechies-4 wavelet decomposition are extracted for texture.

The experiments on Corel dataset show that the combined global and block-based feature representations provide the highest classification accuracy. In addition, increasing the size of training images can provide higher classification accuracy and the region-based feature representation method produces the worst classification performance.

Chen et al [49] propose a neural network model with adaptive structure for image annotation that enables the proposed model to utilize both global and regional visual features. Both a genetic algorithm and the traditional back-propagation algorithm are combined to train the proposed model and it is experimented on a synthetic image dataset. The synthetic image dataset consists of simple objects that can be segmented well using segmentation techniques such as the JSEG algorithm. Six basic colors and nine basic shapes are exploited to construct the synthetic images. The component of the recognition network is a set of sub-networks and the number of sub-networks is determined by the number of segmented regions of an image. During the annotation process, the image is first segmented into several regions. RGB histogram is used to represent the color feature and invariant moments are used to characterize the shape feature. For image annotation, the input layer receive the feature vector X of a region and then the output layer would generate a keyword vector Y to indicate which keyword(s) should be selected to label the input region.

Experiments demonstrate the effectiveness of the proposed model; the correlation network improves annotation performance. However, the synthesis image dataset used is too simple, the method was not experimented on a real image dataset and they did not show how and in which stage the global features are employed.

He et al. [50] propose to extract global features from image local regions (block). They propose to divide the image into some overlapping sub-blocks before applying PCA and 2DPCA (two-dimensional PCA) separately. The traditional PCD/2DPCA are to extract the principal component from global features vectors of a whole image. The authors used them to extract global feature directly from each sub-block, which they call block-global feature. To investigate the performance of the proposed technique, comparative studies have been conducted with SIFT. The experiments were carried out on the visual Object Classed Challenge 2008 and only seven classes were picked up, namely, car, horse, aero plane, bird, cat, chair and sky. SVM is used for learning.

Experiments show that the use of block-global feature is better than the performance of using local feature only, in both single-label and multi-label annotation, and 2DPCA achieves the best and more stable performance in terms of annotation accuracy. However, the performance of PCA is not as good, and is nearly the same with that of SIFT in terms of annotation accuracy.

Kuric [51] and Beilikova and Kuric [52] propose a method to obtain annotation for target images based on a novel combination of local and global feature during search stage. The method consists of two main stages: 1) training dataset preprocessing and 2) processing of target image. The training data set preprocessing consists of image processing, local and global features calculation, plus indexing and clustering of local and global features. The processing of target image consists of the same two previous steps, and querying the keypoint store and global feature index. A disk-based LSH (Locality Sensitive Hashing) approach is employed for indexing and clustering local feature. The feature extraction is performed in three steps: 1) extraction of bounded local features, grid segmentation is used to divide the image to a fixed number of sub-image, 2) extraction of free local features using SIFT and 3) extraction of global feature using Joint Composite Descriptor (JCD) which combines information about colour and texture in a single histogram.

Corel5k was used for the evaluation and the method was compared with the translation model. The proposal method shows better precision than the translation model. However, less recall was noticed compared to the translation model.

VI. Discussion and Conclusions

In this paper, we attempted to provide a comprehensive survey on the latest developments of AIA techniques with a special emphasis on feature extraction. We focused the survey on the emerging image annotation methods, *hybrid methods,* which combine both global and local features for image representation.

In the conventional state of the art of AIA techniques, images are represented by extracted features either locally or globally. A single set of features is computed from the entire image in global methods, which gives compact dimension of the feature vector and makes them relatively less demanding in computational terms. They are simple and easy to extract as no segmentation is needed. Global methods have low computational cost for the feature extraction, they are more distinctive because they have the ability to capture complex and contextual layout information and advantageous in classifying simple scene categories. However, they do not capture spatial information and are weak in characterizing the internal content of image especially when the image has multiple complex objects (see table 1).

Local methods are being increasingly used. Unlike in global methods, images are divided into regions or blocks and a set of features is computed for each of the regions, which means that an image will be represented as a bag of features. A bag of features can represent images at object level and provides spatial information which makes them more precise and discriminating than the global methods. Having said that, local methods pose additional challenges compared to the global methods. Local features may not be accurate due to the usually unsupervised segmentation, and the appearance features extracted from segments are less distinctive and even with perfect segment labels; their union does not always match well. In addition, one image is represented by many visual feature vectors, resulting in high computational cost (see table 1).

TABLE 1
CONTRAST OF THE THREE AIA METHODS

AIA Method	Pros	Cons
Global	Low computational cost for feature extraction [54] Easy to extract and segmentation is not required [49] Compact dimension of the feature vector [52] More distinctive [53] Capture complex information [52] Provide contextual information [48][53] Advantageous in classifying simple scene categories [45] More characteristic layout [53]	Do not capture spatial information [54] Sensitive to intensity variations and distortion [37] Fail to narrow down the semantic gap due to their limited descriptive power based on objects [37] Do not have good performance [54] Weak in characterizing the internal content of image [49] Not recommended for multiple complex object images. Limited interpretability [53]
Local	More precise, discriminating and explicit [53] [52] Spatial information [1] Good for search of specific object [52] [45] Improve classification [48] More flexible and Compositional character [53] Good generalization potential [53]	High computational cost [49] High number of matches for a simple query [52] Need additional processing (e.g. segmentation) Not recommended when searching complex information [52] Produce unsatisfactory accuracy [49]
Hybrid	Combine the advantages of both global and local features [50] More suitable to represent complex scenes and events categories [45] Useful when the choice of one of the global/regional features is not specified [45] Better coherent semantic for annotation Help discovering multiple semantic meanings in one image [50]	Tend to be complex The choice of the feature combinations is not obvious Pose additional challenges May give worse results if the features are not well chosen May requires high dimensional features which imply high computation cost.

All current features have limitations describing images and none of them appears to be powerful enough to represent the large amount and variety of images. Global and local features provide different kinds of information; they have their own advantages in classifying certain categories. However, they have several complementary strengths and there are many situations where the annotation of images should be judged based on the combination of global and local features. The potential for interaction between the two levels is largely unexplored and quite promising.

A potential way forward is to combine the two levels for image representation. This may benefit from the advantages of both global and local features, help discover multiple semantic meanings in one image and improve the annotation performance. However, hybrid methods tend to be complex, and the choice of the feature combinations is not obvious (see table 1). Moreover, the processing and analysis of high dimensional image features is a very complex issue. The performance of existing classifiers degrades considerably when feature dimension is very high. Therefore, features require to be further mined to decide on the right number of features and the right features to be combined for the image representation in order to achieve the annotation task.

These issues represent the subject matter of our future research to supplement the limitations in this area.

REFERENCES

[1] D. Zhang, Md. M. Islam, G. Lu, “A review on automatic image annotation techniques”, Pattern Recognition, vol. 45, no. 1, pp. 346–362, Jan. 2012.

[2] S.K. Chang, A. Hsu, “Image information systems: where do we go from here?” IEEE Transactions on Knowledge and Data Engineering, vol. 5, no. 5, pp. 431–442, 1992.

[3] S. Ornager, “View a picture, theoretical image analysis and empirical user studies on indexing and retrieval”, Swedis Library Research, vol. 2, no. 3, pp. 31-41, 1996.

[4] M. Markkula, E. Sormunen, “End-user searching challenges indexing practices in the digital newspaper photo archive”, Information retrieval, vol. 1, no. 4, pp. 259-285, 2000.

[5] J. Dowe, “Content-based retrieval in multimedia imaging”, in: Proceedings of the SPIE Storage and Retrieval for Image and Video Database, 1993.

[6] J. Laaksonen, M. Koskela, S. Laakso, E.O. Picsom, “Content-based image retrieval with self-organizing maps”, Pattern Recognition Letters, vol. 21, no. 13/14, pp. 1199–1207, 2000.

[7] N. Rasiwasia, N. Vasconcelos, P.J. Moreno, “Query by semantic example”, in: Proceedings of the Fifth International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, Springer, Berlin, vol. 4071, pp. 51–60, 2006.

[8] J. Huang, S.R. Kumar, R. Zabih, “An automatic hierarchical image classification scheme”, in Proceedings of the Sixth ACM International Conference on Multimedia, Bristol, England, pp. 219–228, 1998.

[9] O. Chapelle, P. Haffner, V.N. Vapnik, “Support vector machines for histogram-based image classification”, IEEE Trans. Neural Networks, vol. 10, pp. 1055–1064, 1999.

[10] Y. Gdalyahu, D. Weinshall, “Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouette”, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1312–1328, 1999.

[11] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, H.J. Zhang, “Image classification for content-based indexing”, IEEE Trans. Image Process. vol. 10, no. 1, pp. 117–130, 2001.

[12] A. Yavlinsky, E. Schoeld, S. Ruger, “Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation”,

in Proceedings of the International Conference on Image and Video Retrieval, 2005.
[13] H. Zhang, A. Berg, M. Maire, J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition", In: CVPR (2006).
[14] A. Makadia, V. Pavlovic, S. Kumar, "A new baseline for image annotation", In: ECCV (2008).
[15] B. Babenko, S. Branso, S. Belongie, "Similarity metrics for categorization: from monolithic to category specific", In: ICCV (2009).
[16] M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, "TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation", In: ICCV (2009).
[17] M.M. Gorkani, R.W. Picard, "Texture orientation for sorting photos at a glance", in Proceedings of the 12th International Conference on Pattern Recognition, pp. 459–464, 1994.
[18] M. Szummer, RW. Picard, "Indoor-outdoor image classification", in Proceedings of the IEEE ICCV Workshop on Content-based Access of Image and Video Databases, Bombay, India, pp. 42-51, Jan. 1998.
[19] J. Li, J.Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach", IEEE Trans. PAMI 25, vol. 25, no. 9, pp. 1075–1088, 2003.
[20] Z. Guo, Z.M. Zhang, E.P. Xing, C. Faloutsos, "A Max Margin Framework on image annotation and multimodal image retrieval", in IEEE International Conference on Multimedia and Expo, pp. 504–507, 2007.
[21] N. Serrano, A. Savakis, J. Luo, "A computationally efficient approach to indoor/outdoor scene classification", in Proceedings of the IEEE International Conference on Pattern Recognition, Quebec, Canada, pp. 146-149, 2002.
[22] L. Zhang, J. Ma, "Image annotation by incorporating word correlations into multi-class SVM," Soft Computing - A Fusion of Foundations, Methodologies and Applications, vol. 15, no. 5, pp. 917-927, 2010.
[23] W. Yi, H. Tang, "Experimental analysis on classification of unmanned aerial vehicles images using the Probabilistic Latent Semantic Analysis", International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, Wuhan, China, vol. 7492 no. 3, 2009.
[24] J.R. Smith, C.S. Li, "Image classification and querying using composite region templates", Int. J. Computer Vision Image Understanding, vol. 75, no. 1–2, pp. 165–174, 1999.
[25] K. Barnard, D. Forsyth, "Learning the semantics of words and pictures", in Proceedings of the International Conference on Computer Vision, pp. 408–415, 2001.
[26] P. Duygulu, K. Barnard, J.F.G de Freitas, D.A. Forsiyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary", In the 7th European Conference on Computer vision, Copenhagen, Denmark, pp. 97-112, 2002.
[27] D.M. Blei, M.I. Jordan, "Modeling annotated data", in Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, pp. 127-134, 2003.
[28] C. Yang, M. Dong, F. Fotouhi, "Region based image annotation through multiple instance learning", In Proceedings of the 13th annual ACM International Conference on Multimedia, New York, NY, USA, pp. 435-438, 2005.
[29] J. Tang, P.H Lewis, "Region based image annotation through a training based feature space", submitted to International Conference on Image Processing (ICIP), 2007.
[30] K. Kuroda, M. Hagiwara, "An image retrieval system by impression words and specific object names – IRIS", Neuro-computing, vol. 43, no. 1-4, pp. 259-276, 2002.
[31] A. Hanbury, "A Survey of Methods for Image Annotation", Journal of Visual Languages and Computing, vol. 19, no. 5, pp. 617–627, 2008.
[32] F. Wang, "A Survey on Automatic Image Annotation and Trends of the New Age", Procedia Engineering, vol. 23, pp. 434–438, 2011.
[33] R. Pagare, A. Shinde, "A Study on Image Annotation Techniques", International Journal of Computer Applications, vol. 37, no. 6, pp. 42-45, Jan. 2012.
[34] J. Tang, "Automatic Image Annotation and Object Detection", PhD thesis, School of Electronics and Computer Science, University of Southampton, United Kingdom, May 2008.
[35] R.S. Choras, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", Int. J. of Biology and Biomedical Engineering, vol. 1, no. 1, pp. 6-16, 2007.
[36] E. Saber, A.M. Tekalp, "Integration of color, edge and texture features for automatic region-based image annotation and retrieval", Electronic Imaging, vol. 7, pp. 684–700, 1998.
[37] X. Qi, Y. Han, "Incorporating multiple SVMs for automatic image annotation", Pattern Recognition, Vol. 40, No. 2, pp. 728 – 741, 2007.
[38] D. Kim, H. Park, "An Efficient Face Recognition through Combining Local Features and Statistical Feature Extraction", PRICAI 2010: Trends in Artificial Intelligence, vol. 6230, pp. 456-466, 2010.
[39] Y. Su, S. Shan, X. Chen, W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition", IEEE Transactions on image processing, vol. 18, no. 8, pp.1885 -1896, Aug. 2009.
[40] R.S. Kathavarayan, M. Karuppasamy, "Preserving Global and Local Features for Robust Face Recognition under Various Noisy Environments", International Journal of Image Processing (IJIP), vol. 3, no. 6, pp. 328-340, 2009.
[41] S. Anila, N. Devarajan, "Global and Local Classifiers for Face Recognition", European Journal of Scientific Research, vol. 57, no. 4, pp. 556-566, 2011.
[42] Y. Fang, T. Tan, Y. Wang, "Fusion of Global and Local Features for Face Verification", in Proceeding of ICPR, Vol. 2, pp. 382-385, 2002.
[43] K. Murphy, A. Torralba, D. Eaton, W. Freeman, "Object Detection and Localization Using Local and Global Features, Toward Category-level Object Recognition", Lecture Notes in Computer Science, vol. 4170, pp. 382-400, 2006.
[44] D A. Lisin, M A. Mattar, M B. Blaschko, E G. Learned-Miller, M C. Benfield, "Combining Local and Global Image Features for Object Class Recognition", in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05 Workshops, 2005.
[45] Y. Wang, T. Mei, S. Gonga, X-S. Hua, "Combining global, regional and contextual features for automatic image annotation", Pattern Recognition, vol. 42, no. 2, pp. 259 – 266, 2009.
[46] S. Sarin, W. Kameyama, "Joint Equal Contribution of Global and Local Features for Image Annotation", CLEF working notes 2009, Corfu, Greece, 2009.
[47] C-F. Tsai, W-C. Lin, "A Comparative Study of Global and Local Features Representation in Image Database Categorization", in Fifth International Joint Conference on INC, IMS and IDC, 2009.
[48] J. Ngiam, H. Goh, "Learning Global and Regional Features for Photo Annotation", Multilingual Information Access Evaluation II. Multimedia Experiments, Lecture Notes in Computer Science, vol. 6242, pp. 287 – 290, 2010.
[49] Z. Chen, H. Fu, Z. Chi, D. Feng, "A Neural Network Model with Adaptive Structure for Image Annotation", in 11th Int. Conf. Control, Automation, Robotics and Vision, Singapore, Dec. 2010.
[50] J. He, Z. Jiang, P. Guo, L. Liu, "A Study of Block-Global Feature based Supervised Image Annotation", in IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 971 – 976, Oct. 2011.
[51] E. Kuric, "Automatic Photo Annotation Based on Visual Content Analysis", Information Sciences and Technologies Bulletin of the ACM Slovakia, Special Section on Student Research in Informatics and Information Technologies, vol. 3, no. 2, pp. 72 – 75, 2011.
[52] M. Bieliková, E. Kuric, "Automatic Image Annotation Using Global and Local Features", In Porc. of Semantic Media Adaptation and Personalization, SMAP 2011, Los Alamitos: IEEE Computer Society, pp. 33 – 38, 2011.
[53] D. Kuettel, M. Guillaumin, V. Ferrari, "Combining Image-level and Segment-level Models for Automatic Annotation", In: MMM, Springer, vol. 7131, pp. 16 – 28, 2012.
[54] D-D. Le, S. Satoh, "NII: Japan at ImageCLEF 2011 Photo Annotation Task", CLEF, Notebook Papers/Labs/Workshop, 2011.
[55] C-F. Tsai, C. Hung, "Automatically Annotating Images with Keywords: A Review of Image Annotation Systems", Recent Patents on Computer Science, vol. 1, pp. 55-68, 2008.